

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 10124082 A

(43) Date of publication of application: 15.05.98

(51) Int. Cl

G10L 3/00

G10L 5/04

(21) Application number: 08275791

(71) Applicant: MATSUSHITA ELECTRIC IND CO LTD

(22) Date of filing: 18.10.96

(72) Inventor: NAGAI KIYOTAKA

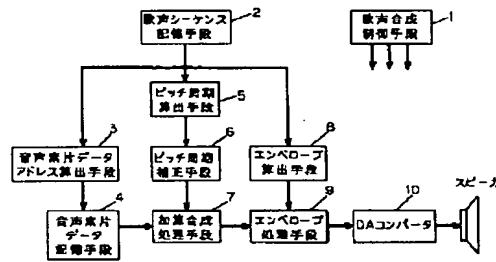
(54) SINGING VOICE SYNTHESIZING DEVICE

(57) Abstract:

PROBLEM TO BE SOLVED: To synthesize a singing voice which is natural and smooth without any feeling of a voice break by the singing voice synthesizing device which uses a PSOLA(pitch synchronous overlap addition of waveform) method.

SOLUTION: Voice waveforms cut in pitch cycles read out of a phoneme piece data storage means 4 are rearranged and put one over another at pitch cycle intervals corresponding to a scale calculated by a pitch cycle calculating means 5 by using an additive synthesizing process means 7 to synthesize a desired singing voice. This device is provided with a pitch cycle correcting means 6 which corrects the pitch cycle intervals so that when voiced sounds are connected to each other, no abrupt variation in pitch cycle interval is caused in the switching timing of the scale, thereby synthesizing the singing voice which is smooth without any feeling of a voice break.

COPYRIGHT: (C)1998,JPO



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-124082

(43)公開日 平成10年(1998)5月15日

(51)Int.Cl.⁶

G 10 L 3/00
5/04

識別記号

F I

G 10 L 3/00
5/04

J
F

審査請求 未請求 請求項の数8 OL (全 11 頁)

(21)出願番号 特願平8-275791

(22)出願日 平成8年(1996)10月18日

(71)出願人 000005821

松下電器産業株式会社
大阪府門真市大字門真1006番地

(72)発明者 永井 清隆
大阪府門真市大字門真1006番地 松下電器
産業株式会社内

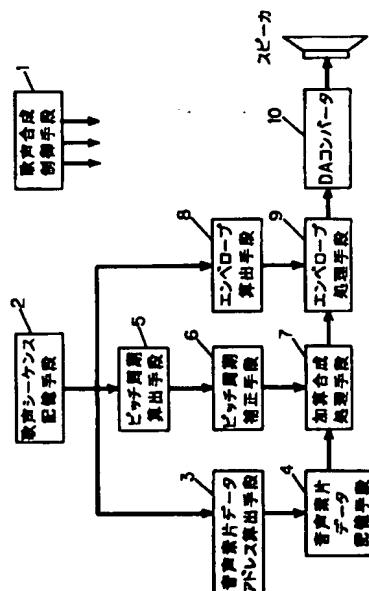
(74)代理人 弁理士 滝本 智之 (外1名)

(54)【発明の名称】 歌声合成装置

(57)【要約】

【課題】 P S O L A (波形のピッチ同期重畠加算) 法を用いた歌声合成装置において、音切れ感がなく自然で滑らかな歌声を合成することを目的とする。

【解決手段】 音声素片データ記憶手段4から読み出したピッチ周期で切り出した音声波形を、ピッチ周期算出手段5で算出した音階に対応するピッチ周期間隔で、加算合成処理手段7を用いて並べ、重ね合わせることにより、所望の歌声を合成する装置であって、有聲音同士を接続する場合、前記音階の切り替えタイミングでピッチ周期間隔の急激な変化が発生しないようにピッチ周期間隔を補正するピッチ周期補正手段6を設けることにより、音切れ感がなく滑らかな歌声を合成することができる。



【特許請求の範囲】

【請求項1】 ピッチ周期で切り出した音声波形を、音階に対応するピッチ周期間隔で並べ、重ね合わせることにより、所望の歌声を合成する装置であって、有声音同士を接続する場合、前記音階の切り替えタイミングでピッチ周期間隔の急激な変化が発生しないように補正して音声波形を並べる手段を設けたことを特徴とする歌声合成装置。

【請求項2】 ピッチ周期で切り出した音声波形を、音階に対応するピッチ周期間隔で並べ、重ね合わせることにより、所望の歌声を合成する装置であって、有声音同士を接続する場合、前記音階の切り替えタイミングでピッチ周期間隔の急激な変化が発生しないように補正して音声波形を並べる手段と、前記音階切り替えタイミング前後の音声波形を前記補正したピッチ周期間隔でオーバーラップして並べ、クロスフェード接続処理を行う手段とを設けたことを特徴とする歌声合成装置。

【請求項3】 音階の開始タイミングと前記音階に対応する歌声の最初の音節の聴覚上のリズム知覚点とをほぼ一致させるように制御して音声波形を並べる手段を設けたことを特徴とする請求項1または2に記載の歌声合成装置。

【請求項4】 音階の開始タイミングと前記音階に対応する歌声の最初の音節の聴覚上のリズム知覚点との時間が所望の時間となるように制御して音声波形を並べる手段を設けたことを特徴とする請求項1または2に記載の歌声合成装置。

【請求項5】 並べた音声波形データのサンプルタイミングが合成音声の出力サンプルタイミングと異なるときには前記サンプルタイミングでの音声波形データを入力として、出力サンプルタイミングでの音声波形データを求める補間手段を設け、補間した音声波形データを並べ、重ね合わせることにより、所望の歌声を合成することを特徴とする請求項1ないし4のいずれかに記載の歌声合成装置。

【請求項6】 無聲音波形データのサンプルタイミングが合成音声の出力サンプルタイミングと一致しない場合、無聲音波形データのサンプルタイミングを出力サンプルタイミングに一致するようにシフトし、無聲音波形データに対しては補間を行わないことを特徴とする請求項5に記載の歌声合成装置。

【請求項7】 補間手段は高次補間を行うことを特徴とする請求項5または6に記載の歌声合成装置。

【請求項8】 音階に対応するピッチ周期間に周期的な変化を与えるピブラート付加手段を設けたことを特徴とする請求項1ないし7のいずれかに記載の歌声合成装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、PSOLA (Pitch-synchronous Overlap-add : 波形のピッチ同期重畠加算)法をもじいて、任意の歌声を合成する歌声合成装置に関する。

【0002】

【従来の技術】 従来、歌声合成装置としては、例えば特開昭62-65098号公報に記載されたものが知られている。前記歌声合成装置は、入力される音声信号をデジタル信号に変換するAD変換器と、このAD変換器によって変換された音声データを記憶するメモリと、このメモリに記憶した音声データを分析し、所用のデータ区間から成る音素を切り出す手段と、この切り出された音素を鍵盤操作によって選択された周期で出力する手段とから構成されている。

【0003】 音素を切り出し、合成するときの具体的なアルゴリズムについては、例えばケイス・レント (Keith Lent) 氏によって、1989年にコンピュータ・ミュージック・ジャーナル (Computer Music Journal) 誌の第13巻、第4号の61頁から75頁に発表された「デジタルサンプル音の効率的なピッチシフト方法」(An Efficient Method for Pitch Shifting Digitally Sampled Sounds) と題する論文に詳細に記載されている。

【0004】 図9に前記論文に記載されている合成処理内容を説明する図を示す。同図に示す全ての図の縦軸は信号振幅を表し、横軸は時間を表す。同図(a)は入力された音声データ、同図(b)は(a)の音声データをピッチ周期毎に切り出すときに用いるピッチ周期の2倍の長さの窓関数を示す。また、同図(c)は(b)の窓関数によって切り出された音声データをピッチ周期間隔で並べ、重ね合わせて、所望のピッチ周期の音声データを合成するときの様子を示している。同図(c)では、入力された音声データのピッチ周期より長いピッチ周期の音声データを合成する場合を示している。

【0005】 このように音声のピッチ周期毎に窓関数(通常、ピッチ周期の2倍程度の長さの窓関数)を用いて音声を切り出し、切り出したピッチ波形を所望のピッチ周期間隔で並べ直し、これらを重ね合わせてピッチ周期を変換した音声を合成する方法をPSOLA法と呼び、従来電子楽器等で用いられてきた時間軸の圧縮・伸張によるピッチ変換法と比較して、フォルマント周波数を大きく変化させることなくピッチ周波数を変化させることができるので、声質の変化が少なく、ピッチ変換範囲の広い音声を合成することができる、という特徴があり、広く用いられてきた。

【0006】

【発明が解決しようとする課題】 しかしながら、従来のPSOLA法を用いた歌声合成装置では、音階の切り替えタイミングで音切れ感が生じ、滑らかな歌声を合成することができない、という問題があった。伴奏音に合わせて歌声を合成する場合、音階の切り替えタイミング、

すなわちピッチ周期の切り替えタイミングは、音符長によって定められる時間で切り替える必要がある。このためピッチ周期の途中でピッチを切り替える場合が生じ、これによりピッチ周期の不連続が生じる。このときの音声波形を図10に示す。同図でT₁とT₂はそれぞれ音階の切り替えタイミングの前と後のピッチ周期を、またT_xは音階を切り替える直前のピッチ波形のピーク値から音階の切り替えタイミングまでの時間を表し、T_xは音階を切り替えるタイミングにより変化する。同図で、

(a) は音階を切り替える前の音声波形を、(b) は音階を切り替えた後の音声波形を、また (c) は、(a) と (b) を接続して合成した音声波形を示す。図10に示す音声波形の場合、音階の切り替えタイミングで (T_x+T₂) という不連続なピッチ周期を生じ、これにより異音が発生する。

【0007】このピッチ周期の不連続による異音の発生を防止するため、従来、図11 (a) に示すフェードアウトした音声波形と同図 (b) に示すフェードインした音声波形を接続して同図 (c) に示す音声波形を合成出力することが行われてきた。これにより、異音の発生を防止できるが、音階の切り替え時に音声波形の振幅減少による音切れ感が生じ、特に有聲音同士を接続する場合にはこの音切れ感によって自然で滑らかな歌声を合成することができないという問題があった。

【0008】また、従来の歌声合成装置では、図7 (a) に示す楽譜に従って、同図 (b) に示すガイドメロディの伴奏音に合わせて歌声を合成する場合、同図 (c) に示したように音階の切り替えタイミングと音階に対応する音節の開始タイミングとを一致させて音声データを接続し、歌声を合成していた。しかしながら、音節の聴覚上のリズム知覚点は必ずしも音節の開始点と一致せず、通常、音節を構成している子音と母音の間にがあるので、リズムを知覚するタイミングが音節の種類によって変動し、合成した歌声のリズム感がよくない、という問題があった。

【0009】本発明は上記問題点を解決するもので、音切れ感がなく自然で滑らかな歌声を合成することができる歌声合成装置を提供することを目的とする。

【0010】また、リズム感がよい歌声を合成することができる歌声合成装置を提供することを目的とする。

【0011】

【課題を解決するための手段】この課題を解決するために、本発明の歌声合成装置は、ピッチ周期で切り出した音声波形を、音階に対応するピッチ周期間隔で並べ、重ね合わせることにより、所望の歌声を合成する装置であって、有聲音同士を接続する場合、前記音階の切り替えタイミングでピッチ周期間隔の急激な変化が発生しないように補正して音声波形を並べる手段を備えている。

【0012】これにより、音切れ感がなく滑らかな歌声を合成する歌声合成装置が得られる。

【0013】また、本発明の歌声合成装置は、音階の切り替えタイミングで切り替え前後の音声波形を前記補正したピッチ周期間隔でオーバーラップして並べ、クロスフェード接続処理を行う手段を備えている。

【0014】これにより、音階の切り替えタイミングで音声波形のスペクトルが徐々に変化し、自然で滑らかな歌声を合成する歌声合成装置が得られる。

【0015】また、本発明の歌声合成装置は、音階の開始タイミングと前記音階に対応する歌声の最初の音節の聴覚上のリズム知覚点とをほぼ一致させるように制御して音声波形を並べる手段を備えている。

【0016】これにより、リズム感がよい歌声を合成する歌声合成装置が得られる。

【0017】

【発明の実施の形態】本発明の第1の発明は、ピッチ周期で切り出した音声波形を、音階に対応するピッチ周期間隔で並べ、重ね合わせることにより、所望の歌声を合成する装置であって、有聲音同士を接続する場合、前記音階の切り替えタイミングでピッチ周期間隔の急激な変化が発生しないように補正して音声波形を並べる手段を設けたものであり、有聲音の音階の切り替えタイミングにおいてもピッチ周期間隔が滑らかに変化するように補正して音声波形を並べ、重ね合わせることにより、ピッチ周期の不連続により発生する異音を防止するため必要であったフェードアウト処理とフェードイン処理を不要とし、音切れ感のない滑らかな歌声を合成することができる。

【0018】第2の発明は、ピッチ周期で切り出した音声波形を、音階に対応するピッチ周期間隔で並べ、重ね合わせることにより、所望の歌声を合成する装置であって、有聲音同士を接続する場合、前記音階の切り替えタイミングでピッチ周期間隔の急激な変化が発生しないように補正して音声波形を並べる手段と、前記音階切り替えタイミング前後の音声波形を前記補正したピッチ周期間隔でオーバーラップして並べ、クロスフェード接続処理を行う手段とを設けたものであり、音階の切り替えタイミングで、切り替え前後の音声波形をクロスフェード処理で接続することにより、音声波形のスペクトルを徐々に変化させ、自然で滑らかな歌声を合成することができる。

【0019】第3の発明は、上記第1または第2の発明に、音階の開始タイミングと前記音階に対応する歌声の最初の音節の聴覚上のリズム知覚点とをほぼ一致させるように制御して音声波形を並べる手段を設けたものであり、これによりリズム感のよい歌声を合成できる。

【0020】第4の発明は、上記第1または第2の発明に、音階の開始タイミングと前記音階に対応する歌声の最初の音節の聴覚上のリズム知覚点との時間が所望の時間となるように制御して音声波形を並べる手段を設けたものであり、これによりリズム感がよく、かつ前記時間

を例えば長めにとることによって、ためて歌う等、歌声としての表現力を増すことができる。

【0021】第5の発明は、上記第1～第4の発明に、並べた音声波形データのサンプルタイミングが合成音声の出力サンプルタイミングと異なるときには前記サンプルタイミングでの音声波形データを入力として、出力サンプルタイミングでの音声波形データを求める補間手段を設け、補間した音声波形データを重ね合わせることにより、所望の歌声を合成するものである。これにより、従来、サンプル周期を単位として表現されていたピッチ周期間隔の精度を例えば12ビット精度分、すなわちサンプル周期の2⁻¹²まで向上させ、自然で滑らかな歌声を合成することができる。

【0022】第6の発明は、上記第5の発明において、無聲音波形データのサンプルタイミングが合成音声の出力サンプルタイミングと一致しない場合、無聲音波形データのサンプルタイミングを出力サンプルタイミングに一致するようにシフトし、無聲音波形データに対しては補間を行わないことにより、高周波数成分を多く含んでいる無聲音の補間にによる音質変化を抑え、自然な歌声を合成することができる。

【0023】第7の発明は、上記第5または第6の発明において、補間手段が高次補間を行うもので、補間精度を向上させることにより、より自然で滑らかな歌声を合成することができる。

【0024】第8の発明は、上記第1～第7の発明に、音階に対応するピッチ周期間隔に周期的な変化を与えるピブラート付加手段を設けたもので、ピッチ周期にピブラートを付加することで、より自然な歌声を合成することができる。

【0025】以下、本発明の実施の形態について、図面を用いて説明する。

(実施の形態1) 図1は、本発明の実施の形態1による歌声合成装置の構成を示すブロック図である。図1において、1は歌声合成制御手段、2は歌声シーケンス記憶手段、3は音声素片データアドレス算出手段、4は音声素片データ記憶手段、5はピッチ周期算出手段、6はピッチ周期補正手段、7は加算合成処理手段、8はエンベロープ算出手段、9はエンベロープ処理手段、10はDAコンバータである。

【0026】以上のように構成された実施の形態1における歌声合成装置について、以下その動作について説明する。

【0027】歌声合成制御手段1は、歌声合成装置全体の動作とそのタイミングを制御する。

【0028】歌声シーケンス記憶手段2は、任意の歌声を合成するのに必要な歌声シーケンスデータを記憶している。

【0029】図2に歌声シーケンス記憶手段2に記憶している歌声シーケンスデータの例を示す。同図におい

て、(a)に示す楽譜に対応する歌声シーケンスデータが、(b)に示されている。同図(b)で音階は音名で表現し、また、発音記号は音節単位で表現している。本実施の形態においては、歌声シーケンスデータは、イベントデータから構成されている。イベントデータは、イベントの発生する時刻と発音記号と音階と音量とから構成され、同図(b)で1行として表されている。イベントデータは、発音記号で指定された音節の音声が、指定された時刻に指定された音階と音量で出力されることを表す。また、音量0は該当する音声の発音終了を表す。

【0030】音声素片データアドレス算出手段3は、歌声シーケンス記憶手段2から読み出された歌声シーケンスデータの発音記号から前記発音記号に対応する音声素片データのアドレスを算出し、音声素片データ記憶手段4に出力する。

【0031】音声素片データ記憶手段4は、ピッチ周期で窓を掛けて切り出した音声波形（音声素片と呼ぶ）データを記憶しており、音声素片データアドレス算出手段3から出力されたアドレスにしたがって音声素片データを読み出し、出力する。

【0032】一方、ピッチ周期算出手段5は、歌声シーケンス記憶手段2から読み出された歌声シーケンスデータの音階から前記音階に対応するピッチ周期間隔を算出する。

【0033】次に、ピッチ周期補正手段6は、音階の切り替えタイミングで有聲音同士を接続する場合、ピッチ周期間隔の急激な変化が生じないように、ピッチ周期算出手段5から出力されたピッチ周期間隔を補正し、出力する。

【0034】ピッチ周期補正手段6の動作について図面を用いてさらに説明する。音階の切り替えタイミングでの補正前のピッチ周期を図10に示すものとしたとき、ピッチ周期補正手段6によって補正されたピッチ周期を図3に示す。図3は、実施の形態1において、(a)は音階を切り替える前の音声波形を、(b)は音階を切り替えた後の音声波形を、(c)は(a)と(b)を接続して合成した音声波形を示す。音階の切り替えタイミングで、図10に示す補正前のピッチ周期は($T_x + T_2$)であるのに対し、図3に示す補正後のピッチ周期は($T_1 + T_2$)／2である。すなわち、ピッチ周期補正手段6は、音階の切り替えタイミングでピッチ周期の急激な変化が発生することを検出し、この実施の形態では切り替え前後のピッチ周期の平均値に置き換える。

【0035】加算合成処理手段7は、音声素片データ記憶手段4から読み出された音声素片データをピッチ周期補正手段6で補正されたピッチ周期間隔で並べ、重ね合わせて図3(c)に示す音声を合成し、出力する。

【0036】エンベロープ算出手段8は、歌声シーケンス記憶手段2から読み出された歌声シーケンスデータの音量から前記音量に対応するゲインを算出し、エンベロ

ープ情報（この場合はゲインのみ）をエンベロープ処理手段9に出力する。

【0037】エンベロープ処理手段9は、エンベロープ算出手段8からのエンベロープ情報にしたがって加算合成手段7から出力された合成音声に対してエンベロープ処理を行い、結果の歌声合成音声をDAコンバータ10を介してスピーカから出力する。

【0038】以上のように、実施の形態1によれば、有声音同士を接続する場合、前記音階の切り替えタイミングでピッチ周期間隔の急激な変化が発生しないように補正する手段を設けることにより、有聲音の音階の切り替えタイミングにおいてもピッチ周期間隔を滑らかに変化させ、音階の切り替えタイミングでのフェードアウト処理とフェードイン処理を不要とすることによって、音切れ感のない滑らかな歌声を合成することができる。

【0039】なお、実施の形態1では、音声素片データ記憶手段4は、ピッチ周期で窓を掛けて切り出した音声波形を記憶しているものとしたが、切り出し前の音声波形データを記憶し、読み出すときに窓を掛けることによって切り出してもよい。

【0040】（実施の形態2）図4は、本発明の実施の形態2による歌声合成装置の構成を示すブロック図である。図4において、1は歌声合成制御手段、2は歌声シーケンス記憶手段、3は音声素片データアドレス算出手段、4は音声素片データ記憶手段、5はピッチ周期算出手段、6はピッチ周期補正手段、7は加算合成処理手段、8はエンベロープ算出手段、9はエンベロープ処理手段、10はDAコンバータ、11は加算合成処理手段である。

【0041】実施の形態1と実施の形態2の構成上の相違点は、実施の形態2では加算合成処理手段11が追加されている点である。

【0042】以上のように構成された実施の形態2における歌声合成装置について、前記構成上の相違点を中心にして、以下その動作について説明する。

【0043】歌声合成制御手段1は、歌声合成装置全体の動作とそのタイミングを制御する。歌声シーケンス記憶手段2は、任意の歌声を合成するのに必要な歌声シーケンスデータを記憶している。歌声シーケンス記憶手段2に記憶している歌声シーケンスデータは、実施の形態1で説明したものと同一である。音声素片データアドレス算出手段3は、歌声シーケンス記憶手段2から読み出された歌声シーケンスデータの発音記号から前記発音記号に対応する音声素片データのアドレスを算出し、音声素片データ記憶手段4に記憶しており、音声素片データアドレス算出手段3から出力されるアドレスにしたがって音声素片データを読み出し、出力する。

【0044】一方、ピッチ周期算出手段5は、歌声シーケンス記憶手段2から読み出された歌声シーケンスデータの音階から前記音階に対応するピッチ周期間隔を算出し、出力する。次に、ピッチ周期補正手段6は、音階の切り替えタイミングで有聲音同士を接続する場合、実施の形態1で説明したのと同様にして、ピッチ周期間隔の急激な変化が生じないように補正したピッチ周期間隔を加算合成手段7と加算合成処理手段11に出力する。

ケンス記憶手段2から読み出された歌声シーケンスデータの音階から前記音階に対応するピッチ周期間隔を算出し、出力する。次に、ピッチ周期補正手段6は、音階の切り替えタイミングで有聲音同士を接続する場合、実施の形態1で説明したのと同様にして、ピッチ周期間隔の急激な変化が生じないように補正したピッチ周期間隔を加算合成手段7と加算合成処理手段11に出力する。

【0045】加算合成処理手段7は、音声素片データ記憶手段4から読み出された音声素片データをピッチ周期補正手段6で補正されたピッチ周期間隔で並べ、音階単位で重ね合わせて合成音声を出力する。実施の形態1では、加算合成処理手段7は音階間の重ね合わせ（加算合成）も行っていたのに対して、実施の形態2では、加算合成処理手段7は音階単位での重ね合わせを行い、音階間の重ね合わせは加算合成手段11で行う。

【0046】次に、エンベロープ算出手段8の動作について、図5を用いて説明する。エンベロープ算出手段8は、図5(a)に示すように、音階の切り替え前の音声波形をピッチ周期補正手段6で補正したピッチ間隔で並べた音声波形に対してフェードアウト処理を行うのに必要なエンベロープ情報（図5の破線で示す）を算出する。前記エンベロープ情報の最大振幅は、実施の形態1と同様に、歌声シーケンスデータの音量に対応するゲインにより定められる。さらに、エンベロープ算出手段8は、図5(b)に示すように、音階の切り替え後の音声波形をピッチ周期補正手段6で補正したピッチ間隔で並べた音声波形に対してフェードイン処理を行うのに必要なエンベロープ情報も算出する。図5に示したように、フェードアウトする音声波形とフェードインする音声波形は、音節の切り替えタイミングをほぼ中心にして互いにオーバーラップしている。エンベロープ算出手段8は、以上のようにして算出したエンベロープ情報をエンベロープ処理手段9に出力する。

【0047】エンベロープ処理手段9は、エンベロープ算出手段8からのエンベロープ情報にしたがって加算合成手段7から出力された音階単位の合成音声に対してエンベロープ処理を行い、図5(a), (b)に示す合成音声波形を出力する。

【0048】加算合成処理手段11は、エンベロープ処理手段9から出力された音階単位の合成音声をオーバーラップして接続し、加算合成処理を行うことにより、図5(c)に示す音声を合成し、出力する。

【0049】次に、加算合成処理手段11から出力された歌声合成音声をDAコンバータ10を介してスピーカから出力する。

【0050】以上のように、実施の形態2によれば、有聲音同士を接続する場合、音階の切り替えタイミングでピッチ周期間隔の急激な変化が発生しないように補正する手段と、前記音階切り替えタイミング前後の音声波形

を前記補正したピッチ周期間隔でオーバーラップして並べ、クロスフェード接続処理を行う手段とを設けることにより、音階の切り替えタイミングにおいて音声波形のスペクトルを徐々に変化させることができるので、自然で滑らかな歌声を合成することができる。

【0051】(実施の形態3) 図6は、本発明の実施の形態3による歌声合成装置の構成を示すブロック図である。図6において、1は歌声合成制御手段、2は歌声シーケンス記憶手段、3は音声素片データアドレス算出手段、4は音声素片データ記憶手段、5はピッチ周期算出手段、6はピッチ周期補正手段、7は加算合成処理手段、8はエンベロープ算出手段、9はエンベロープ処理手段、10はDAコンバータ、11は加算合成処理手段、12は音節開始時刻算出手段、13はビブラート付加手段、14はサンプル間データ補間手段である。

【0052】実施の形態2と実施の形態3の構成上の相違点は、実施の形態3では音節開始時刻算出手段12、ビブラート付加手段13、サンプル間データ補間手段14が追加されている点である。

【0053】以上のように構成された実施の形態3における歌声合成装置について、前記構成上の相違点を中心にして、以下その動作について説明する。

【0054】歌声合成制御手段1は、歌声合成装置全体の動作とそのタイミングを制御する。歌声シーケンス記憶手段2は、任意の歌声を合成するのに必要な歌声シーケンスデータを記憶している。歌声シーケンス記憶手段2に記憶している歌声シーケンスデータは、実施の形態1で説明したものと同一である。

【0055】音節開始時刻算出手段12は、歌声シーケンス記憶手段2から読み出された歌声シーケンスデータの音階の開始時刻と発音記号とから、音階の開始時刻と前記音階に対応する最初の音節の聴覚上のリズム知覚点とがほぼ一致するように音節の開始時刻を算出する。

【0056】以下、図7を用いて音節開始時刻算出手段12の動作についてさらに説明する。図7において、

(a)は以下の説明で例として用いる楽譜、(b)は前記楽譜の第1小節に示されたメロディの音階(音名で表記)とその時間長(単位はms)、(c)は従来の音階の開始時刻と前記音階に対応する最初の音節の開始時刻の時間関係、(d)は実施の形態3における音階の開始時刻と前記音階に対応する最初の音節の開始時刻の時間関係を表す。同図(d)で▽は音節のリズム知覚点を表す。

【0057】音節のリズム知覚点とは、聴覚上のリズム知覚の基準点のことと、通常、音節を構成している子音と母音との間にある。同図(c)と(d)においては、音節の発音記号を子音と母音とに分けて表し、(d)では音節のリズム知覚点の位置を模式的に示している。

【0058】音節の聴覚上のリズム知覚点の算出法については、例えば、新居康彦、大崎正巳著の「音声処理と

DSP」(1989年啓学出版発行)と題する本の164頁から166頁に記載されている。なお、この本では、音節のリズム知覚点のことを拍同期点、あるいは受聴タイミング点と呼んで説明している。

【0059】従来、図7(c)に示すように音階の開始時刻と音節の開始時刻とを一致させていたのに対し、本実施の形態では、同図(d)に示すように、音階の開始時刻と音節のリズム知覚点とを一致させる。実際には、有聲音同士を接続する場合、音階の切り替えタイミングでピッチ周期が不連続とならないように制御しているので音階の開始時刻と音節のリズム知覚点とが若干ずれる場合がある。

【0060】音節開始時刻算出手段12は、各音節の音節開始点からリズム知覚点までの時間を予め記憶しており、音階の開始時刻から前記リズム知覚点までの時間分逆のぼった時刻を音節の開始時刻として算出し、出力する。

【0061】音声素片データアドレス算出手段3は、音節開始時刻算出手段12で算出された音節の開始時刻を用いて前記音節に対応する音声素片データのアドレスを算出し、音声素片データ記憶手段4に出力する。

【0062】音声素片データ記憶手段4は、ピッチ周期で切り出した音声素片波形を記憶しており、音声素片データアドレス算出手段3から出力されるアドレスにしたがって音声素片データを読み出し、出力する。

【0063】一方、ピッチ周期算出手段5は、歌声シーケンス記憶手段2から読み出された歌声シーケンスデータの音階から前記音階に対応するピッチ周期間隔を算出し、出力する。次に、ピッチ周期補正手段6は、音階の切り替えタイミングで有聲音同士を接続する場合、実施の形態1で説明したのと同様にして、ピッチ周期間隔の急激な変化が生じないように補正したピッチ周期間隔を出力する。

【0064】ビブラート付加手段13は、ピッチ周期補正手段6から出力された補正したピッチ周期間隔に周期的な変動を与え、ビブラートを付加し、ビブラートを付加したピッチ周期間隔をサンプル間データ補間手段14と加算合成処理手段7と加算合成処理手段11とに出力する。ピッチ周期にビブラートを付加することで、より自然な歌声を合成することができる。

【0065】以下、サンプル間データ補間手段14の動作について図8を用いて説明する。同図で縦の実線が音声素片データ記憶手段4に記憶されたサンプルタイミングでのデータ値であり、縦の破線が出力サンプルタイミングにおけるデータ値である。また、同図でTは音声素片データのサンプル周期を表す。同図は、ピッチ周期間隔が8.5Tで2つの音声素片データ(同図(a))、

(b)に示す)を並べ、重ね合わせて合成する(同図(c)に示す)場合を示す。音声素片データ記憶手段4から読み出された2つの音声素片波形データの内、

(a) では音声素片データ記憶手段4から読み出された音声素片データのサンプルタイミングと出力タイミングが一致している。同図(a)で破線は見えていないが、これは実線と破線が重なっているためである。一方、同図(b)では音声素片データ記憶手段4から読み出された音声素片データのサンプルタイミングと出力タイミングが0.5T分ずれている。

【0066】サンプル間データ補間手段14は、音声素片データ記憶手段4から読み出されたサンプルタイミングの音声素片波形データから出力サンプルタイミングにおける音声素片波形データを補間して出力する。出力サンプルタイミングの前後のサンプルタイミングの音声素片波形データから出力サンプルタイミングにおける音声素片波形データを直線補間することにより求め、出力する。すなわち、図8(b)に示すように、縦の実線で示す音声素片波形データを入力として破線で示す出力サンプルタイミングの音声素片波形データを補間して算出し、出力する。

【0067】なお、無聲音波形データに対しては、前記無聲音波形データのサンプルタイミングが合成音声の出力サンプルタイミングと一致しない場合、前記無聲音波形データのサンプルタイミングを出力サンプルタイミングに一致するようにシフトし、補間を行わないようにすることにより、高周波数成分を多く含んでいる無聲音の補間にによる音質変化を抑え、より自然な歌声を合成することができる。

【0068】加算合成処理手段7は、サンプル間データ補間手段14から出力された音声素片データをピラート付加手段13から出力されたピッチ周期間隔で並べ、音階単位で重ね合わせて合成音声を出力する。

【0069】エンベロープ算出手段8は、実施の形態2で説明したのと同様に、有聲音同士を接続する場合、音階の切り替えタイミング前後の音声波形を補正したピッチ周期間隔でオーバーラップして並べ、クロスフェード接続処理を行うのに必要なエンベロープ情報を算出し、エンベロープ処理手段9に出力する。

【0070】エンベロープ処理手段9は、エンベロープ算出手段8からのエンベロープ情報にしたがって加算合成手段7から出力された音階単位での合成音声に対してエンベロープ処理を行う。加算合成処理手段11は、エンベロープ処理手段7から出力された音階単位の合成音声をオーバーラップして接続、加算合成処理を行うことで歌声合成音声を出力する。次に、加算処理手段11から出力される歌声合成音声をDAコンバータ10を介してスピーカから出力する。

【0071】以上のように実施の形態3によれば、音階の開始タイミングと前記音階に対応する歌声の最初の音節の聴覚上のリズム知覚点とをほぼ一致させるように前記音節の開始時刻を制御する手段を設けることにより、リズム感のよい歌声を合成することができる。

【0072】なお、音階の開始タイミングと前記音階に対応する歌声の最初の音節の聴覚上のリズム知覚点との時間が所望の時間となるように前記音節の開始時刻を制御する手段を設け、例えば前記所望の時間を長めにとることによって、ためて歌う感じを出す等、歌声としての表現力を増すことができる。

【0073】また、実施の形態3においては、サンプル間データ補間手段14で直線補間を行うことにしたが、必ずしも直線補間でなく高次の補間演算を行うようにしても良い。これにより補間精度を向上させ、より自然で滑らかな歌声を合成することができる。

【0074】

【発明の効果】以上のように本発明は、ピッチ周期で切り出した音声波形を、音階に対応するピッチ周期間隔で並べ、重ね合わせることにより、所望の歌声を合成する装置であって、有聲音同士を接続する場合、音階の切り替えタイミングでピッチ周期間隔の急激な変化が発生しないように補正して音声波形を並べる手段を設けたものであり、有聲音の音階の切り替えタイミングにおいてもピッチ周期間隔を滑らかに変化させ、音階の切り替えタイミングでのフェードアウト処理とフェードイン処理を不要とすることによって、音切れ感のない滑らかな歌声を合成することができる。

【0075】また、本発明は、ピッチ周期で切り出した音声波形を、音階に対応するピッチ周期間隔で並べ、重ね合わせることにより、所望の歌声を合成する装置であって、有聲音同士を接続する場合、音階の切り替えタイミングでピッチ周期間隔の急激な変化が発生しないよう補正して音声波形を並べる手段と、音階切り替えタイミング前後の音声波形を補正したピッチ周期間隔でオーバーラップして並べ、クロスフェード接続処理を行う手段とを設けたものであり、音階の切り替えタイミングにおいて音声波形のスペクトルを徐々に変化させることにより、自然で滑らかな歌声を合成することができる。

【0076】また、本発明は、音階の開始タイミングと音階に対応する歌声の最初の音節の聴覚上のリズム知覚点とをほぼ一致させるように制御して音声波形を並べる手段を設けたものであり、これによりリズム感のよい歌声を合成することができる。

【図面の簡単な説明】

【図1】本発明の実施の形態1による歌声合成装置の構成を示すブロック図

【図2】同歌声合成装置の歌声シーケンス記憶手段に記憶している歌声シーケンスデータの例を示す図

【図3】同歌声合成装置の音階切り替えタイミングにおける歌声合成波形を説明するための図

【図4】本発明の実施の形態2による歌声合成装置の構成を示すブロック図

【図5】同歌声合成装置の音階切り替えタイミングにおける歌声合成波形を説明するための図

【図6】本発明の実施の形態3による歌声合成装置の構成を示すブロック図

【図7】同歌声合成装置の音階の開始タイミングと音節の開始タイミングとの時間関係を示す図

【図8】同歌声合成装置のサンプル間データ補間手段の動作を説明するための図

【図9】従来の音声合成方法の処理内容を説明するための図

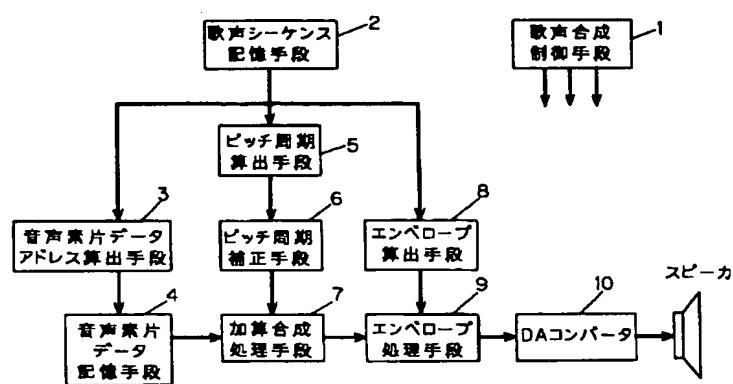
【図10】音階の切り替えタイミングでピッチ周期の不連続が発生することを説明するための図

【図11】従来の歌声合成装置の音階の切り替えタイミングにおける歌声合成波形を説明するための図

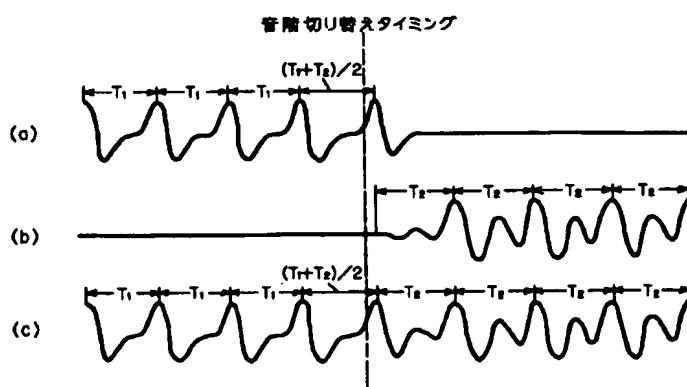
【符号の説明】

- 1 歌声合成制御手段
- 2 歌声シーケンス記憶手段
- 3 音声素片データアドレス算出手段
- 4 音声素片データ記憶手段
- 5 ピッチ周期算出手段
- 6 ピッチ周期補正手段
- 7, 11 加算合成処理手段
- 8 エンベロープ算出手段
- 9 エンベロープ処理手段
- 10 10 DAコンバータ
- 12 音節開始時刻算出手段
- 13 ピブラー付加手段
- 14 サンプル間データ補間手段

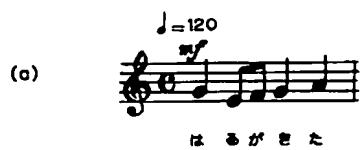
【図1】



【図3】



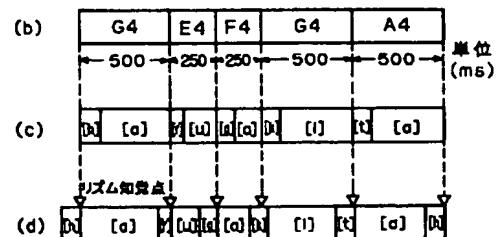
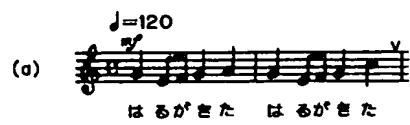
【図2】



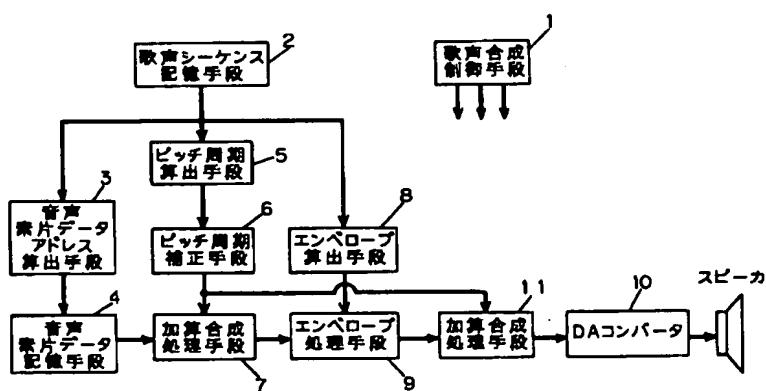
(b)

時刻	発音記号	音階	音量
t_0	[ho]	G4	6
t_1	[ho]	G4	0
t_2	[ru]	E4	6
t_3	[ru]	E4	0
t_4	[ga]	F4	6
t_5	[ga]	F4	0
t_6	[ki]	G4	6
t_7	[ki]	G4	0
t_8	[ta]	A4	6
t_9	[ta]	A4	0

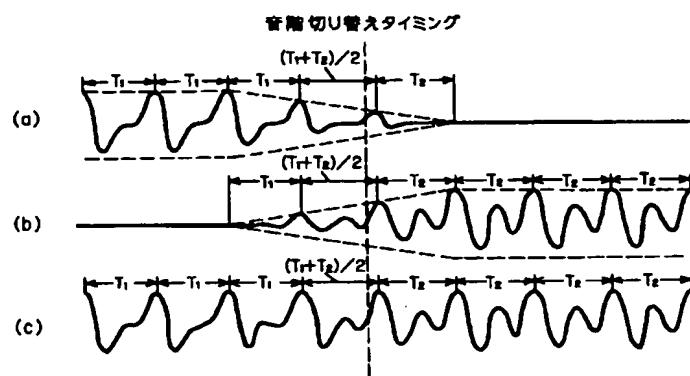
【図7】



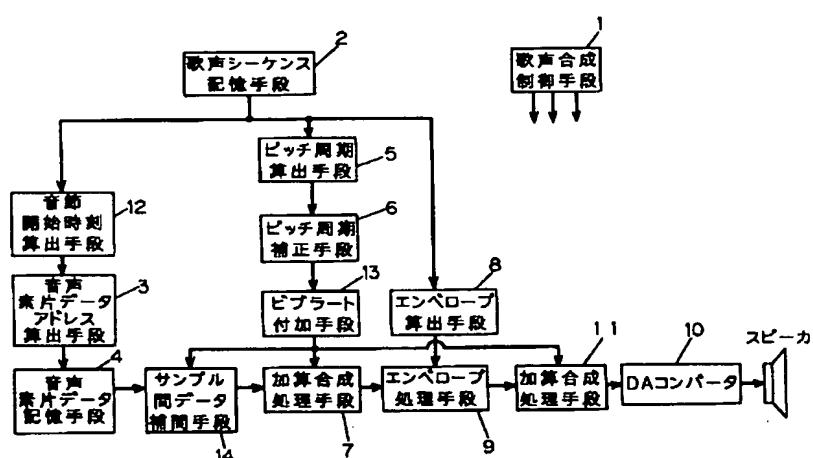
【図4】



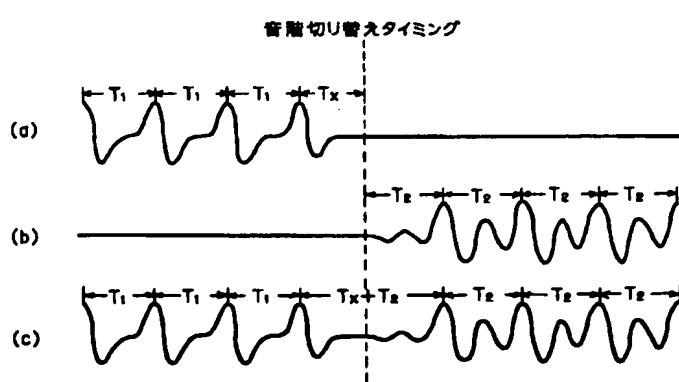
【図5】



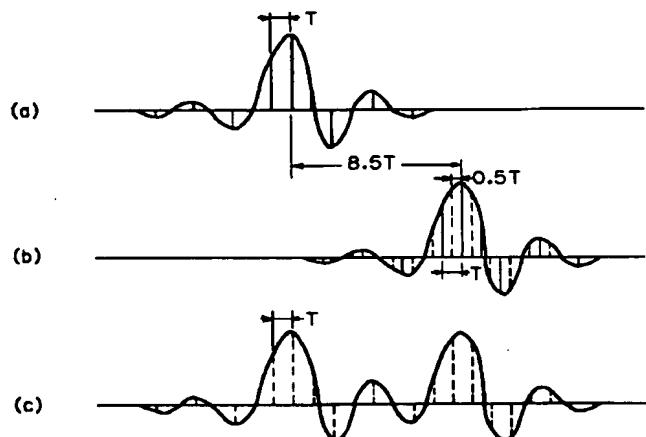
【図6】



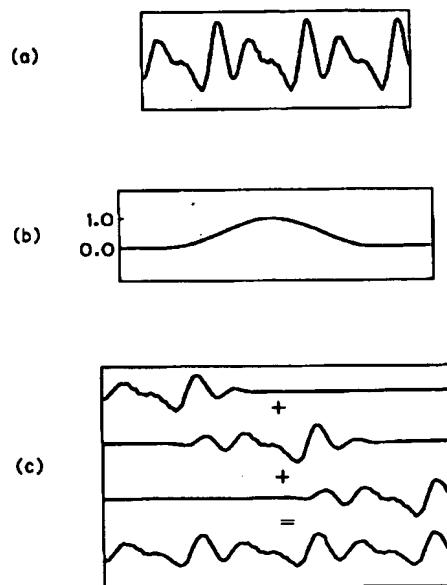
【図10】



【図8】



【図9】



【図11】

